

ABSTRACT

A language model generation and accumulation apparatus that generates and accumulates language models for speech recognition. The language model generation and accumulation
5 apparatus is comprised of: a higher-level N-gram generation and accumulation unit that generates and accumulates a higher-level N-gram language model obtained by modeling each of a plurality of texts as a string of words including a word string class having a specific linguistic property; and a lower class dependent word
10 N-gram generation and accumulation unit that generates and accumulates a lower-level N-gram language model obtained by modeling a sequence of words included in each word string class.